

MULTISCALE INTEGRATORS FOR STOCHASTIC DIFFERENTIAL EQUATIONS AND IRREVERSIBLE LANGEVIN SAMPLERS

JIANFENG LU AND KONSTANTINOS SPILIOPOULOS

ABSTRACT. We study multiscale integrator numerical schemes for a class of stiff stochastic differential equations (SDEs). We consider multiscale SDEs that behave as diffusions on graphs as the stiffness parameter goes to its limit. Classical numerical discretization schemes, such as the Euler-Maruyama scheme, become unstable as the stiffness parameter converges to its limit and appropriate multiscale integrators can correct for this. We rigorously establish the convergence of the numerical method to the related diffusion on graph, identifying the appropriate choice of discretization parameters. Theoretical results are supplemented by numerical studies on the problem of the recently developing area of introducing irreversibility in Langevin samplers in order to accelerate convergence to equilibrium.

1. INTRODUCTION

The main focus of this work is numerical integrators for stochastic differential equations (SDEs) with multiscale coefficients. SDEs have been used ubiquitously in many areas to model dynamical systems under random perturbations due to e.g., extrinsic noises caused by thermal fluctuations, the degrees of freedom that are not explicitly tracked by the system, or perhaps some intrinsic randomness of the system. Due to their wide applications, numerical solution to SDEs has been an active area of research in numerical analysis and applied probability.

While our results can be used in other scenarios, the main motivation of this work is to design numerical integrators for SDEs arising from recent works on irreversible Langevin sampling [19, 21, 22]. In those works, an irreversible drift term is added to the overdamped Langevin dynamics (details will be specified in section 2), and it is proved that under the proper assumptions the sampling efficiency increases as the magnitude of the irreversible drift goes to infinity, which is validated by numerical studies, see [7, 21, 22]. However, at the same time, when a large irreversible drift is added to the original overdamped Langevin equation, the stiffness of the system is inevitably increased, and thus prevents the application of standard numerical integrators to the resulting systems. The goal of this work is to study multiscale integrators that allow to enlarge the magnitude of the irreversible drift without having to sacrifice the stability of the numerical algorithm.

For SDEs with multiscale coefficients, it is well understood that we shall take into account the multiscale structure in order to design better integrators (see e.g., the books [8, 20]). The key idea is to use the averaged limit of the SDE when the scale is well separated. Hence, it is not necessary to accurately resolve the scales of the original system, but we can rather work with the averaged limit. This has been the underlying principle of the heterogeneous multiscale methods (HMM) [3, 9, 10], in particular see [11, 26, 27] for its applications to stochastic differential equations. Other numerical approaches for stiff SDEs were also developed in [1, 2, 4, 6, 18, 25].

Date: July 8, 2016.

J.L. was partially supported by the National Science Foundation under the grant DMS-1415939. K.S. was partially supported by the National Science Foundation (NSF) DMS 1312124 and by NSF CAREER award DMS 1550918.

Our numerical scheme follows the ideas of the FLAVORS method developed in [25], which on the algorithmic level is very similar to the seamless version of HMM method developed in [12, 13]. The basic idea is to use a split-step integrator which combines a short time integration of the whole SDE and a longer time integration of the SDE without the stiff terms. The numerical analysis of such schemes [25] shows that in the case that the variables of the SDEs can be one-to-one mapped to a set of “fast” and “slow” variables, the numerical scheme converges to the averaged limit which consists of the dynamics of the slow component. We emphasize that the algorithm does not require explicit knowledge of the mapping that transforms the system into fast and slow variables, while it does require the forcing terms of the SDE can be separated into stiff and non-stiff terms.

The main contribution of this work is to extend the analysis of [25] to situations that a one-to-one mapping of the original degree of freedom into fast and slow variables is not possible. In particular, for the irreversible Langevin sampling the function U that maps the configurational space to the energy is clearly not one-to-one. In fact, it is well known that in the limit the SDE converges to a diffusion on an associated graph [15, 16], for which besides the energy, one has to add the index variable to represent the state space. Our main result proves that the multiscale integrator converges to a diffusion on graph as the scale separation parameter tends to infinity and the discretization parameters are appropriately chosen. In the one well case, our proof follows ideas of [25] appropriately adjusting for the different limiting behavior that we have here. Then, the results are being extended to the multiple well case by using techniques similar to those of the classical averaging techniques of [5, 14, 16]. However, since we work in the discrete time framework and not in the continuous time framework, we need to obtain bounds with explicit dependence on the discretization parameters.

This paper is organized as follows. We will introduce the SDEs from the irreversible Langevin sampler and the HMM multiscale integrator in Section 2. Some numerical results are presented in Section 3 to validate the method. The averaging results of the SDEs, in particular, convergence to the diffusion on graphs are recalled in Section 4. The main results and the proofs are given in Section 5.

2. HMM INTEGRATOR FOR IRREVERSIBLE LANGEVIN SAMPLING SCHEME

Consider the overdamped Langevin equation

$$(1) \quad dZ_t = -\nabla U(Z_t) dt + \sqrt{2\beta} dW_t,$$

where $U : E \rightarrow \mathbb{R}$ is a given potential, $\beta = (k_B T)^{-1}$ is the inverse temperature, and W_t is the standard multi-dimensional Wiener process. See Section 4 for conditions on U . The overdamped Langevin dynamics (1) is often used to sample the Boltzmann-Gibbs measure, given by

$$\varrho(z) \propto e^{-U(z)/\beta},$$

which is the invariant measure of (1) under mild conditions. Note that the infinitesimal generator of (1) is self-adjoint with respect to the invariant measure, and thus the dynamics (1) is reversible in time, *i.e.*, it satisfies detailed balance.

In [19, 21, 22], it was proposed to add to the overdamped Langevin dynamics an irreversible forcing to accelerate the sampling, the resulting dynamics reads

$$(2) \quad dZ_t^\varepsilon = \left[-\nabla U(Z_t^\varepsilon) + \frac{1}{\varepsilon} C(Z_t^\varepsilon) \right] dt + \sqrt{2\beta} dW_t$$

with some initial condition $Z_0 = z_0$. The invariant measure is maintained if the vector fields C satisfies $\text{div}(Ce^{-U/\beta}) = 0$, or equivalently

$$\text{div } C = \beta^{-1} C \nabla U.$$

A convenient choice, which we assume henceforth, is to pick C such that

$$\text{div } C = 0, \quad \text{and} \quad C \nabla U = 0.$$

This is not the most general choice for C , but it has the advantage that allows to choose C independently of β . One such choice of $C(z)$ is $C(z) = J \nabla U(z)$, where J is any antisymmetric matrix. These conditions mean that the flow generated by C preserves Lebesgue measure since it is divergence-free, at the same time, since U is a constant of the motion, the micro-canonical measure on the surfaces $\{U = z\}$ are preserved as well.

The amplitude of the irreversible drift in (2) is chosen to be $\frac{1}{\varepsilon}$. We will consider the regime that $\varepsilon \ll 1$. Using the large deviation action functional of the empirical measure, it is shown in [21, 22] that the dynamics (2) converges faster to the invariant measure for a larger irreversible drift, i.e., as ε becomes smaller. From another point of view, as will be recalled in section 4, in the limit $\varepsilon \rightarrow 0$, the slow component associated to the solution of the SDE (2) converges to the averaging limit which is a diffusion on an associated graph, and hence the entropy associated with the iso-surfaces of U is completely removed and only the energetic barrier is left in the limit.¹ In [22] it is also established that the asymptotic (as $t \rightarrow \infty$) variance of the estimator is decreasing in ε and in the limit as $\varepsilon \downarrow 0$, it converges to the asymptotic (as $t \rightarrow \infty$) variance of the corresponding sampling problem on the graph where the limiting diffusion lives.

Increasing the irreversible drift however comes with a price: The right hand side of the SDE (2) becomes rather stiff as $\varepsilon \rightarrow 0$, and as a result, standard integrators (for example the Euler-Maruyama scheme) would require vanishingly small time step size to resolve the fast scale of the dynamics. As ε goes to zero, the SDE contains multiple time scale, and thus it is better to use multiscale integrators for such dynamics.

In this work, we investigate a multiscale integrator for stiff SDEs as (2) proposed in [25], which is also rather close to the seamless version of HMM scheme [11–13]. For a macro time step δ and micro time step τ such that $\tau \ll \varepsilon \ll \delta$, from t_n to $t_n + \delta$, we evolve the dynamics

$$(3a) \quad d\bar{Z}_t = [-\nabla U(\bar{Z}_t) + \frac{1}{\varepsilon} C(\bar{Z}_t)] dt + \sqrt{2\beta} dW_t \quad t \in [t_n, t_n + \tau);$$

$$(3b) \quad d\bar{Z}_t = -\nabla U(\bar{Z}_t) dt + \sqrt{2\beta} dW_t \quad t \in [t_n + \tau, t_n + \delta).$$

This can be understood as a split-step time integrator where for the short time step τ we use the whole SDE and for the long time step $\delta - \tau$ we neglect the irreversible drift. The equations above can be integrated using standard numerical schemes, and for definiteness, in this work we will discretize using the standard Euler-Maruyama method, which gives

$$(4a) \quad \bar{Z}_{t_n+\tau} - \bar{Z}_{t_n} = -\tau \nabla U(\bar{Z}_{t_n}) + \frac{\tau}{\varepsilon} C(\bar{Z}_{t_n}) + \sqrt{2\beta\tau} \xi_n;$$

$$(4b) \quad \bar{Z}_{t_n+\delta} - \bar{Z}_{t_n+\tau} = -(\delta - \tau) \nabla U(\bar{Z}_{t_n+\tau}) + \sqrt{2\beta(\delta - \tau)} \xi'_n,$$

where ξ_n and ξ'_n are independent standard normal random variables.

¹While it is possible to combine the irreversible sampling with other techniques to overcome the energetic barrier, we will not go further in this direction as it is not the focus of the current work.

Remark. As will be discussed in Sections 4 and 5, if the dynamical system $\dot{z}_t = C(z_t)$ does not have a unique invariant measure on each connected component of the level sets of $U(z)$ and the dimension is bigger than two, then one needs to modify the scheme by considering an additional regularizing noise, see (8)-(9).

We will show that with proper choices of the time steps τ and δ as $\varepsilon \rightarrow 0$, the numerical scheme (4) converges to the diffusion on graphs, which is the averaging limit of (2). Thus, we may use (4) to numerically discretize the SDE which is consistent in the asymptotic regime as $\varepsilon \rightarrow 0$.

Let us remark that it is also possible to use multiple micro steps with length τ rather than just one such step as in (4), which is analogous to the original HMM integrators. For the purpose of sampling invariant measure, one could also combine the integrator with Metropolis adjustment steps as in the MALA method [23]. We will focus on the numerical analysis of the scheme (4) and leave these extensions to future works.

3. NUMERICAL EXAMPLES

Before we turn to the analytical results, let us present a few numerical tests for the multiscale integrator. We will first consider the sampling efficiency of the irreversible Langevin sampler with the multiscale HMM integrator. We will then show some numerical examples illustrating properties of the integrator. We limit ourselves to simple toy examples as the focus is to demonstrate the numerical properties of the integrator and validate the numerical analysis results, rather than applying to scheme to realistic problems.

For the first test, we consider a 2D symmetric double well potential given by

$$U(x, y) = \frac{1}{4}(x^2 - 1)^2 + y^2,$$

with inverse temperature $\beta = 0.1$. We choose $C = J\nabla U$ with $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. The initial condition is set to be the origin, and we consider the empirical average of the observable $f(x, y) = x + y^2$ over the time interval $[0, 2000]$ with a burn-in period $T_{\text{burn}} = 20$. To test the performance of the sampling scheme based on the multiscale integrator, we compare the empirical average with the true average of the observable with respect to the invariant measure. Note that in this case, due to the choice of the potential and the observable, the true average is explicitly given by $\langle f \rangle = \frac{1}{2}\beta = 0.05$. We also estimate the asymptotic variance of the sampling scheme by dividing the sampled data points into 20 batches.

Denote the sampling error (with respect to the observable f) as Err_f and the asymptotic variance (with respect to the observable f) as AVar_f . The numerical results for various choice of ε are shown in Table 1, in which we also include the results for direct Euler-Maruyama discretization for comparison. In these tests, we fix the macro time step $\delta = 5e-3$ (for the direct Euler-Maruyama discretization, δ is the time step size), and choose the micro time step $\tau = 10\delta\varepsilon = 0.05\varepsilon$. For a given set of parameters, we report the mean and standard deviation estimated from 2000 independent runs of the algorithms. We note that the Euler-Maruyama scheme is unstable for ε below 0.1 with the given δ .

We remark that the particular choice of the observable $f = x + y^2$ makes the accurate sampling rather challenging in this case: As $\mathbb{E}x = 0$ due to the symmetry, the correct sampling of the average value requires fine balance of the time the trajectory spent in left and right well of the double well potential. This explains the high relative error that $\mathbb{E}(\text{Err}_f)$ is on the same order of $\mathbb{E}f$.

We make several observations in regards to the numerical results in Table 1. First, from the result of the Euler-Maruyama scheme for various ε , it is clear that a larger irreversible drift (smaller ε)

	ε	τ	δ	$\mathbb{E}(\text{Err}_f)$	$\text{Std}(\text{Err}_f)$	$\mathbb{E}(\text{AVar}_f)$	$\text{Std}(\text{AVar}_f)$
E-M	5		5e-3	1.2478e-1	9.3383e-2	3.4198e-1	7.5015e-2
	5e-1		5e-3	7.1563e-2	5.4549e-2	1.4521e-1	4.2309e-1
	1e-1		5e-3	4.0312e-2	2.5798e-2	1.7275e-2	5.5163e-2
HMM	1e-2	5e-4	5e-3	3.9982e-2	2.5348e-2	1.7397e-2	5.4576e-2
	1e-3	5e-5	5e-3	3.9529e-2	2.5706e-2	1.7107e-2	5.4546e-2
	1e-4	5e-6	5e-3	3.9090e-2	2.4613e-2	1.7301e-2	5.5807e-2

TABLE 1. Comparison of the Euler-Maruyama scheme and the HMM multiscale integrator for the double well potential (3). The macro time step is $\delta = 5e-3$ with the micro time step $\tau = 0.05\varepsilon$. The mean and standard deviation of the sampling error Err_f and asymptotic variance AVar_f are reported for various choice of ε . The Euler-Maruyama scheme is not stable for smaller value of ε (for example, $\varepsilon = 5e-2$) and thus not reported.

	ε	τ	δ	$\mathbb{E}(\text{Err}_f)$	$\text{Std}(\text{Err}_f)$	$\mathbb{E}(\text{AVar}_f)$	$\text{Std}(\text{AVar}_f)$
E-M	5e-2		1e-3	3.2342e-2	2.2662e-2	1.6678e-2	5.3673e-3
HMM	1e-3	2e-5	1e-3	3.1832e-2	2.2346e-2	1.6515e-2	5.3137e-3
	1e-4	2e-6	1e-3	3.1628e-2	2.2233e-2	1.6462e-2	5.3621e-3
	1e-5	2e-7	1e-3	3.1228e-2	2.2487e-2	1.6524e-2	5.2017e-3
HMM	1e-4	1e-6	1e-3	3.6570e-2	2.7574e-2	3.7407e-2	1.1820e-2
	1e-5	1e-7	1e-3	3.5446e-2	2.6375e-2	3.7668e-2	1.1787e-2

TABLE 2. Comparison of the Euler-Maruyama scheme and the HMM multiscale integrator for the double well potential (3) with reduced macro time step $\delta = 1e-3$ (compared with $\delta = 5e-3$ in Table 1). The Euler-Maruyama scheme is now stable with $\varepsilon = 5e-2$, but not stable for smaller values of ε . The micro time step in HMM scheme is chosen to be either $\tau = 20\delta\varepsilon = 0.02\varepsilon$ or $\tau = 10\delta\varepsilon = 0.01\varepsilon$. The mean and standard deviation of the sampling error Err_f and asymptotic variance AVar_f are reported for various choice of ε .

enhances the sampling as the sampling error and also the asymptotic variance decrease. Second, while for a fixed computational cost the Euler-Maruyama scheme becomes unstable for small ε , the HMM scheme works well for smaller ε which further reduces the sampling error. Moreover, we remark that while the integrator works well for very small ε , the improvement in this example for going to a very small ε is limited, this is expected since when $\varepsilon \rightarrow 0$, as will be shown later, the scheme becomes an approximation of the averaging limit of the SDE. Thus the sampling efficiency is determined by the limiting system, and the impact of a finite but small ε may be negligible. Of course, this depends on how fast ergodicity kicks in allowing the averaging limit to be achieved. Note that the HMM multiscale integrator allowed us to reach to the $\varepsilon \rightarrow 0$ limit stably, while the Euler-Maruyama scheme blows up for small values of ε keeping δ fixed.

We further test the dependence of the HMM scheme on the choice of parameters in Table 2. In those tests, we decrease the value of macro time step to $\delta = 1e-3$. In comparison, we also list the result of the Euler-Maruyama scheme with the same time step, which is now stable for smaller $\varepsilon = 5e-2$ (but loses stability if we further reduce ε). The results for the HMM scheme with different

ε and $\tau = 20\delta\varepsilon = 0.02\varepsilon$ suggest that it is better to take a smaller ε , though the improvement is again marginal in this case. Compared with Table 1, we see that a smaller δ improves the sampling results, though of course this comes with a higher computational cost.

In Table 2, we also consider choice of the micro time step τ with different ratios of $\tau/(\delta\varepsilon)$ to see the dependence. We observe that in the case with the smaller δ , if we still take τ such that $\tau = 10\delta\varepsilon$ as in Table 1, the performance of the sampling scheme is in fact worse than the direct Euler-Maruyama scheme (with a larger ε). Thus it motivates the choice of a larger τ , which increases the effective sampling time of the fast dynamics, and hence is expected to lead to better performance. This is confirmed in the numerical results with the choice of $\tau = 20\delta\varepsilon = 0.02\varepsilon$. Unfortunately, if we further increase τ (choosing for example $\tau = 30\delta\varepsilon = 0.03\varepsilon$ here), the HMM scheme becomes unstable. This instability can be understood in our theoretical analysis as the assumption that $(\tau/\varepsilon)^{3/2} \ll \delta$ in the convergence result Theorem 6. With a fixed macro time step δ (and hence fixing computational budget), it seems that a good practice is to choose a larger ratio $\tau/(\varepsilon\delta)$ while making sure that the scheme is stable.

In the next example, we consider a more complicated potential, still in two dimension, given by

$$U(x, y) = \frac{1}{4} \left[(x^2 - 1)^2 ((y^2 - 2)^2 + 1) + 2y^2 - y/8 \right] + e^{-8x^2 - 4y^2}$$

with inverse temperature $\beta = 0.2$. This is the potential considered in [22, Example 3]. We take the observable $f(x, y) = (x - 1)^2 + y^2$ and setting $\delta = 5e-3$ and $\tau = 10\delta\varepsilon = 0.05\varepsilon$. The total simulation time is $T = 2000$ with a burn-in period $T_{\text{burn}} = 20$. 2000 independent runs of the algorithms are used to get statistics of the sampling error and asymptotic variance. The results are reported in Table 3. Here the true average of the observable is obtained by a discretization of the Gibbs distribution on the phase space with a fine mesh, which gives approximately $\mathbb{E}f \approx 2.1986$. Similarly as in the double well potential the Euler-Maruyama scheme loses stability for ε smaller than 0.1. The conclusion of the numerical results is similar to that of the double well example.

	ε	τ	δ	$\mathbb{E}(\text{Err}_f)$	$\text{Std}(\text{Err}_f)$	$\mathbb{E}(\text{AVar}_f)$	$\text{Std}(\text{AVar}_f)$
E-M	5		5e-3	5.0964e-1	3.6448e-1	2.4957e00	5.3972e-1
	5e-1		5e-3	3.1799e-1	2.3650e-1	1.8792e00	3.7341e-1
	1e-1		5e-3	1.0730e-1	8.0109e-2	3.4238e-1	1.0624e-1
HMM	1e-2	5e-4	5e-3	1.0347e-1	7.6486e-2	3.0648e-1	9.2405e-2
	1e-3	5e-5	5e-3	1.0255e-1	7.7384e-2	2.9778e-1	9.1258e-2
	1e-4	5e-6	5e-3	1.0108e-1	7.7460e-2	2.9760e-1	8.8149e-2

TABLE 3. Comparison of the Euler-Maruyama scheme and the HMM multiscale integrator for the potential (3). The macro time step is $\delta = 5e-3$ and the micro time step $\tau = 0.05\varepsilon$. The mean and standard deviation of the sampling error Err_f and asymptotic variance AVar_f are reported for various choice of ε .

Next we plot the x coordinate of a sample trajectory and the corresponding potential energy $U(x(t), y(t))$ for the double well potential in Figure 1. The simulation is done with $\varepsilon = 1e-5$ and time step sizes $\tau = 5e-7$ and $\delta = 0.05$. The trajectory is plotted at the end of every macro step (so on the interval of δ). The plot focuses on the time period $[20, 21]$ during which the trajectory mainly stays in the left portion of the phase space $\{x < 0\}$. As can be clearly observed from the figure, while the solution of the SDE oscillates very fast, the potential energy U changes much more slowly, which suggests that U is a slow variable in the averaging limit.

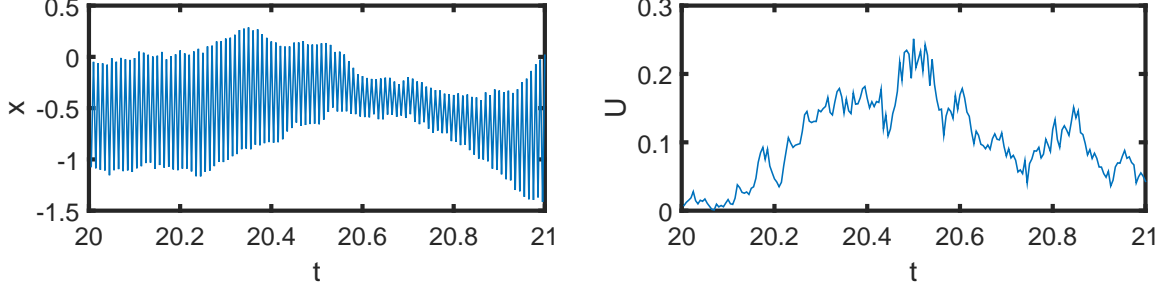


FIGURE 1. A sample trajectory of the HMM multiscale integrator for the double well potential (3) with $\varepsilon = 1e-5$, $\tau = 5e-7$, and $\delta = 0.05$. (Left) The x coordinate of the trajectory for $t \in [20, 21]$. (Right) The potential energy U associated with the trajectory.

It is clear that the map from (x, y) to U is not one-to-one in these examples. For the double well potential, below the energy of the saddle point $U(0, 0) = \frac{1}{4}$, the isopotential curve is disjoint and separated into the left and right half-planes corresponding to the two minima of the potential at $(-1, 0)$ and $(1, 0)$, which are represented on the corresponding graph of the potential as two separated edges that meet at the interior vertex $U = \frac{1}{4}$ corresponding to the saddle point (see section 4 where these concepts are recalled). Therefore, to get from one component of the isopotential curve to the other, the trajectory has to go up in energy and cross the interior vertex; when the energy is decreased from above $\frac{1}{4}$, the trajectory would go into one of the edges, corresponding to one of the disjoint components. In fact, such an event of energy goes above the saddle point energy can be observed already in Figure 1 around $t = 20.5$, where the energy first goes up and when it drops down, the trajectory goes back to the same potential well (on the left half-plane).

To see how the multiscale integrator captures diffusion across the interior vertices on the graph, we record the number of transitions to each edge with lower energy connecting to the saddle point when the energy of the trajectory is decreasing from above that of the vertex. For the double well potential, we count the transitions into each component for a long trajectory with total time $T = 2000$ (with burn-in time $T_{\text{burn}} = 20$) when the energy decreases from above $\frac{1}{4}$. The simulation parameters are $\varepsilon = 1e-5$, $\tau = 5e-7$ and $\delta = 0.005$. For a single realization of the algorithm, we obtain

$$N_{\text{left}} = 3719, \quad \text{and} \quad N_{\text{right}} = 3659,$$

where N_{left} denotes the number of times the trajectory goes to the left well and N_{right} for the right well during the time period $[T_{\text{burn}}, T]$. Note that the empirical probability of going to the left well is 0.5041, very close to the theoretical value of the diffusion on the graphs (which is 0.5 due to the symmetry). While the data reported is only for one realization, this is the typical behavior observed for the algorithm.

In comparison, let us now consider a similar test for a tilted double well potential (so that the symmetry is broken):

$$U(x, y) = \frac{1}{4}(x^2 - 1)^2 - \frac{1}{8}x + y^2.$$

The tilting by $-\frac{1}{8}x$ moves the local minima and the saddle point of the potential to approximately $(-0.9304, 0)$, $(1.0575, 0)$, and $(-0.12705, 0)$. We repeat the same calculation as in the symmetric double well case and obtain

$$N_{\text{left}} = 2485, \quad \text{and} \quad N_{\text{right}} = 3704,$$

so that the empirical probability of going to the left well is 0.4015. Due to the asymmetry, it is less likely to go to the left well when the energy decreases from above the value of the saddle point. This is consistent with the theoretical results for the multiple well case that we will establish in section 5.

4. THE AVERAGING PROBLEM

The irreversible perturbations with a small ϵ induce a fast motion on the constant potential surface and slow motion in the orthogonal direction. Using the theory of diffusions on graphs and the related averaging principle, see [5, 14, 16], we may identify the limiting motion of the slow component, see [22]. The fast motion on constant potential surfaces decreases the variance of the estimator as the phase space is explored more efficiently. Let us consider the level set

$$d(x) = \{z \in E : U(z) = x\},$$

where E denotes the state space. We then denote by $d_i(x)$ the connected components of $d(x)$, i.e.,

$$d(x) = \bigcup_i d_i(x).$$

We define Γ to be the graph which is homeomorphic to the set of connected components $d_i(x)$ of the level sets $d(x)$. Exterior vertexes correspond to minima of U , whereas interior vertexes correspond to saddle points of U . The edges of Γ are indexed by I_1, \dots, I_m . Each point on Γ is indexed by a pair $y = (x, i)$ where x is the value of U on the level set corresponding to y and i is the edge number containing y . Clearly the pair $y = (x, i)$ forms a global coordinate on Γ . Let $Q : E \mapsto \Gamma$ with $Q(z) = (U(z), i(z))$ be the corresponding projection on the graph. For an edge I_k and a vertex O_j we write $I_k \sim O_j$ if O_j lies at the boundary of the edge I_k . We endow the tree Γ with the natural topology. It is known that Γ forms a graph with interior vertexes of order two or three, see for example [17].

If the dynamical system $\dot{z}_t = C(z_t)$ does not have a unique invariant measure on each connected component $d_i(x)$, then we may need to regularize the problem by introducing an additional artificial noise component in the fast dynamics, i.e.,

$$(5) \quad dZ_t^\epsilon = \left[-\nabla U(Z_t^\epsilon) dt + \sqrt{2\beta} dW_t \right] + \left[\frac{1}{\epsilon} \tilde{C}(Z_t^\epsilon) dt + \sqrt{\frac{\kappa}{\epsilon}} \sigma(Z_t^\epsilon) dW_t^o \right].$$

Here, W and W^o are independent standard Wiener processes, the matrix σ will be specified below, and we have defined

$$\tilde{C}_i(z) = C_i(z) + \frac{\kappa}{2} \sum_{j=1}^d \frac{\partial [\sigma \sigma^T(z)]_{j,i}}{\partial z_j}, \quad i = 1, \dots, d.$$

If $\kappa = 0$ then the fast motion is the deterministic dynamical system $\dot{z}_t = C(z_t)$ and Z_t^ϵ is a random perturbation of this dynamical system. For example, if d is even we can take C to be the Hamiltonian vector field $C(z) = J \nabla U(z)$. If $\kappa > 0$ we have random perturbations of diffusion processes with a conservation law.

We make several technical assumptions on $C(z)$, $U(z)$ and $\sigma(z)$ in order to guarantee that the averaging principle applies to (5). We make these assumptions in order to guarantee that the fast process has a unique invariant measure and will have U as a smooth first integral. In order to guarantee the existence of a unique invariant measure for the fast dynamics we assume:

Condition 1. *In dimension $d = 2$, we take $\kappa \geq 0$. In dimension $d > 2$, we either assume that the dynamical system $\dot{z}_t = C(z_t)$ has a unique invariant measure on each connected component $d_i(x)$, in which case we take $\kappa \geq 0$, or otherwise we assume that $\kappa > 0$.*

As far as the potential function $U(x)$ and the perturbation $C(x)$ are concerned, we shall assume:

Condition 2. *The potential function $U(x)$ and the perturbation $C(x)$ satisfy*

- (1) *There exists $a > 0$ such that $U \in \mathcal{C}^{(2+a)}(E)$ and $C \in \mathcal{C}^{(1+a)}(E)$.*
- (2) *$\operatorname{div} C(z) = 0$ and $C(z) \cdot \nabla U(z) = 0$.*
- (3) *U has a finite number of critical points z_1, \dots, z_m and at these points the Hessian matrix D^2U is non-degenerate.*
- (4) *There is at most one critical point for each connected level set component of U .*
- (5) *If z_k is a critical point of U , then there exists a constant $d_k > 0$ such that $C(z) \leq d_k |z - z_k|$.*
- (6) *If $d = 2$ and $\kappa = 0$, then $C(z) = 0$ implies $\nabla U(z) = 0$ and for any saddle point z_k of $U(z)$, there exists a constant $c_k > 0$ such that $|C(z)| \geq c_k |z - z_k|$.*

In regards to the additional artificial perturbation by the noise W_t° , i.e., when $\kappa > 0$, we assume:

Condition 3. (1) *The matrix $\sigma(z)\sigma^T(z)$ is symmetric, non-negative definite, and with smooth entries.*

- (2) *$\sigma(z)\sigma^T(z)\nabla U(z) = 0$ for all $z \in E$.*
- (3) *For any $z \in E$ and $\xi \in \mathbb{R}^d$ such that $\xi \cdot \nabla U(z) = 0$ we have that $\lambda_1(z)|\xi|^2 \leq \xi^T \sigma(z)\sigma^T(z)\xi \leq \lambda_2(z)|\xi|^2$ where $\lambda_1(z) > 0$ if $\nabla U(z) \neq 0$ and there exists a constant K such that $\lambda_2(z) < K$ for all $z \in E$. Moreover if z_k is a critical point for U , then there are positive constants k_1, k_2 such that for all z in a neighborhood of z_k*

$$\lambda_1(z) \geq k_1 |z - z_k|^2, \text{ and } \lambda_2(z) \leq k_2 |z - z_k|^2.$$

- (4) *Let $\lambda_{i,k}$ be the eigenvalues of the Hessian of $U(z)$ at the critical points z_k where $k = 1, \dots, m$ and $i = 1, \dots, d$. Then we assume that $\kappa < (K \max_{i,k} \lambda_{i,k})^{-1}$.*

We remark here that the end result does not depend on the additional regularizing noise, since $\sigma(z)$ does not appear in the limiting dynamics. Let us next identify the corresponding fast and slow components. The fast motion corresponds to the infinitesimal generator

$$(6) \quad \hat{\mathcal{L}}g(z) = \tilde{C}(z)\nabla g(z) + \frac{\kappa}{2} \operatorname{tr} [\sigma \sigma^T(z) D^2g(z)] = C(z)\nabla g(z) + \frac{\kappa}{2} \nabla [\sigma \sigma^T(z) \nabla g(z)] .$$

Let us write \hat{Z}_t for the diffusion process that has infinitesimal generator $\hat{\mathcal{L}}$. Conditions 2 and 3 guarantee that with probability one, if the initial point of \hat{Z} is in a connected component $d_i(x)$, then $\hat{Z}_t \in d_i(x)$ for all $t \geq 0$. Indeed, by Itô formula we have

$$U(\hat{Z}_t) = U(\hat{Z}_0) + \int_0^t \hat{\mathcal{L}}U(\hat{Z}_s) ds + \int_0^t \nabla U(\hat{Z}_s) \sigma(\hat{Z}_s) dW_s.$$

Since $C(z)\nabla U(z) = 0$ and $\sigma(z)\sigma^T(z)\nabla U(z) = 0$ we obtain with probability one $\int_0^t \hat{\mathcal{L}}U(\hat{Z}_s) ds = 0$. The quadratic variation of the stochastic integral is also zero, due to $\sigma(z)\sigma^T(z)\nabla U(z) = 0$, which implies that with probability one $\int_0^t \nabla U(\hat{Z}_s) \sigma(\hat{Z}_s) dW_s = 0$. Thus, we indeed get that for all $t \geq 0$ $\hat{Z}_t \in d_i(x)$ given that the initial point belongs to the particular connected component of the level

set $d_i(x)$. In particular, Itô formula gives for a test function $f \in \mathcal{C}^2(\mathbb{R})$

$$\begin{aligned} f(U(Z_t^\varepsilon)) &= f(U(z)) + \int_0^t \left(-|\nabla U(Z_s^\varepsilon)|^2 + \beta \operatorname{tr} [D^2 U(Z_s^\varepsilon)] \right) f'(U(Z_s^\varepsilon)) + \beta |\nabla U(Z_s^\varepsilon)|^2 f''(U(Z_s^\varepsilon)) ds \\ &\quad + \sqrt{2\beta} \int_0^t f'(U(Z_s^\varepsilon)) \nabla U(Z_s^\varepsilon) dW_s \end{aligned}$$

where Z^ε is the solution to (5).

Let $m(z)$ be a smooth invariant density with respect to Lebesgue measure for the process \widehat{Z}_t . Then, the proof of [14, Lemma 2.3] and the fact that $t \geq 0$ $\widehat{Z}_t \in d_i(x)$ if $\widehat{Z}_0 \in d_i(x)$ imply that if $(x, i) \in \Gamma$ is not a vertex, there exists a unique invariant measure $\mu_{x,i}$ concentrated on the connected component $d_i(x)$ of $d(x)$ which takes the form

$$\mu_{x,i}(A) = \frac{1}{T_i(x)} \oint_A \frac{m(z)}{|\nabla U(z)|} \ell(dz),$$

where $\ell(dz)$ is the surface measure on $d_i(x)$ and $T_i(x) = \oint_{d_i(x)} \frac{m(z)}{|\nabla U(z)|} \ell(dz)$. Notice that if $(x, i) \in \Gamma$ is not a vertex, then the invariant density on $d_i(x)$ is

$$m_{x,i}(z) = \frac{m(z)}{T_i(x) |\nabla U(z)|}, \quad z \in d_i(x).$$

We remark here that in the case $\kappa > 0$, it is relatively easy to see that independently of the form of the matrix $\sigma(z)\sigma^T(z)$, the fact that $\operatorname{div}(C) = 0$ implies that the Lebesgue measure is invariant for the diffusion process corresponding to the operator $\widehat{\mathcal{L}}$. Hence, in that case any constant function is an invariant density. Also, in the case $d = 2$ and $\kappa = 0$, one immediately obtains from Condition 2 that $m(z) = \frac{|\nabla U(z)|}{|C(z)|}$, see [14, Proposition 2.1].

Given a sufficiently smooth function $f(z)$, define its average over the related connected component of the level set of $U(z)$ by

$$\widehat{f}(x, i) = \oint_{d_i(x)} f(z) m_{x,i}(z) \ell(dz) = \frac{1}{T_i(x)} \oint_{d_i(x)} \frac{f(z)}{|\nabla U(z)|} m(z) \ell(dz)$$

We write \mathcal{L}_0 for the infinitesimal generator of the process Z_t given by (2) with $C(z) = 0$. Let us then set

$$\begin{aligned} \widehat{\mathcal{L}_0 U}(x, i) &= \oint_{d_i(x)} \mathcal{L}_0 U(z) m_{x,i}(z) \ell(dz) = \frac{1}{T_i(x)} \oint_{d_i(x)} \frac{\mathcal{L}_0 U(z)}{|\nabla U(z)|} m(z) \ell(dz), \\ \widehat{A}(x, i) &= \oint_{d_i(x)} 2\beta |\nabla U(z)|^2 m_{x,i}(z) \ell(dz) = \frac{1}{T_i(x)} \oint_{d_i(x)} \frac{2\beta \nabla U(z) \cdot \nabla U(z)}{|\nabla U(z)|} m(z) \ell(dz) \end{aligned}$$

and then consider the one-dimensional process Y_t on the branch I_i governed by the infinitesimal generator

$$(7) \quad \mathcal{L}_i^Y g(x) = \widehat{\mathcal{L}_0 U}(x, i) g'(x) + \frac{1}{2} \widehat{A}(x, i) g''(x)$$

Within each edge I_i of Γ , $Q(Z_t^\varepsilon) = (U(Z_t^\varepsilon), i(Z_t^\varepsilon))$ converges as $\varepsilon \downarrow 0$ to a process with infinitesimal generator \mathcal{L}_i^Y . In order to uniquely define the limiting process, we need to specify the behavior at the vertexes of the tree, which amounts to imposing restrictions on the domain of definition of the generator, denoted by \mathcal{L}^Y , of the Markov process. For this purpose, we have the following definition

Definition 4. We say that g belongs in the domain of definition of \mathcal{L}^Y , denoted by $\mathcal{D}(\mathcal{L}^Y)$, of the diffusion Y , if

- (1) The function $g(x)$ is twice continuously differentiable in the interior of an edge I_i .

- (2) The function $x \mapsto \mathcal{L}_i^Y g(x)$ is continuous on Γ .
 (3) At each interior vertex O_j with edges I_i that meet at O_j , the following gluing condition holds

$$\sum_{i: I_i \sim O_j} \pm b_{ji} D_i g(O_j) = 0$$

where, if γ_{ji} is the separatrixes curves that meet at O_j , we have set

$$b_{ji} = \oint_{\gamma_{ji}} \frac{2\beta |\nabla U(z)|^2}{|\nabla U(z)|} m(z) \ell(dz) = 2\beta \oint_{\gamma_{ji}} |\nabla U(z)| m(z) \ell(dz).$$

Here one chooses $+$ or $-$ depending on whether the value of U increases or decreases respectively along the edge I_i as we approach O_j . Moreover D_i represents the derivative in the direction of the edge I_i .

Moreover, within each edge I_i the process Y_t is a diffusion process with infinitesimal generator \mathcal{L}_i^Y .

Consider now the process Y_t that has the aforementioned \mathcal{L}^Y as its infinitesimal generator with domain of definition $\mathcal{D}(\mathcal{L}^Y)$, as defined in Definition 4. Such a process is a continuous strong Markov process, e.g., [17, Chapter 8]. Then, for any $T > 0$, $Q(Z_t^\varepsilon)$ converges weakly in $\mathcal{C}([0, T]; \Gamma)$ to the Markov process Y_t on the tree as $\varepsilon \downarrow 0$. In particular, we have the following theorem.

Theorem 5 (Theorem 2.1 of [14]). *Let Z_t^ε be the process that satisfies (5). Assume Conditions 1, 2 and 3. Let $T > 0$ and consider the Markov process on the tree $\{Y_t, t \in [0, T]\}$ as defined by Definition 4. We have*

$$Q(Z^\varepsilon) \rightarrow Y, \text{ weakly in } \mathcal{C}([0, T]; \Gamma), \text{ as } \varepsilon \downarrow 0.$$

5. ANALYSIS OF THE NUMERICAL HMM METHOD

To analyze the numerical scheme, following [25], let us assume that there exists a random variable $\Phi_h^\alpha(z)$ and an h_0 such that for all $0 < h \leq h_0$ and $\alpha = 0$ or $\frac{1}{\varepsilon}$ one has the estimate

$$(8) \quad \left(\mathbb{E} \left| \Phi_h^\alpha(z) - z + h \nabla U(z) - \alpha h \tilde{C}(z) - \sqrt{h} \sqrt{2\beta} \xi - \sqrt{h} \sqrt{\kappa \alpha} \sigma(z) \xi' \right|^2 \right)^{1/2} \leq C h^{3/2} (1 + \alpha)^{3/2},$$

where ξ, ξ' are independent standard normal random variable. Then, the algorithm becomes

$$(9) \quad \begin{aligned} \bar{Z}_0^\varepsilon &= z_0 \\ \bar{Z}_{(\kappa+1)\delta}^\varepsilon &= \left(\Phi_{\delta-\tau}^0 \circ \Phi_\tau^{\frac{1}{\varepsilon}} \right) (\bar{Z}_{\kappa\delta}^\varepsilon) \end{aligned}$$

Recall from Theorem 5 that it is important to keep in mind that Z_t^ε does not converge to somewhere when $\varepsilon \downarrow 0$. What converges to somewhere, i.e., to the diffusion on the tree, is $Q(Z^\varepsilon) = (U(Z^\varepsilon), i(Z^\varepsilon))$. Then, we have the following theorem.

Theorem 6. *Assume the conditions of Theorem 5 and that $\varepsilon, \delta, \tau \downarrow 0$ are such that $\frac{\delta\varepsilon}{\tau}, \frac{\tau}{\varepsilon}, \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta} \downarrow 0$. Then, for $\tau < \delta < \frac{\tau}{\varepsilon} \ll 1$ sufficiently small, the process $Q(\bar{Z}_{n\delta}^\varepsilon) = (U(\bar{Z}_{n\delta}^\varepsilon), i(\bar{Z}_{n\delta}^\varepsilon))$ (where \bar{Z}^ε is the process from (9)) converges in distribution to the process Y as defined in Definition 4. In addition, convergence to the invariant measure μ of the Y process holds, in the sense that for any bounded and uniformly Lipschitz test function f we have that for all $t > 0$*

$$\lim_{h \downarrow 0} \lim_{\varepsilon, \delta, \frac{\delta\varepsilon}{\tau}, \frac{\tau}{\varepsilon}, \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta} \downarrow 0} \frac{1}{h} \int_t^{t+h} E_\pi f(\bar{Z}_s^\varepsilon) ds = E_\mu \hat{f}(Y_t),$$

where π is the invariant measure of the continuous process Z^ε .

5.1. The case of one well. Let us assume that there is only one well, i.e., that for any $T > 0$ and $s \in [0, T]$, we have $i(\bar{Z}_s^\varepsilon) = 1$. In this case, we simply have $Q(\bar{Z}_s^\varepsilon) = (U(\bar{Z}_s^\varepsilon), 1)$ and we are interested in the asymptotic behavior of the process $Q_s = U(\bar{Z}_s^\varepsilon)$. Going back to (8) we have the following lemma.

Lemma 1. *Consider z such that $|\nabla U(z)| \leq C < \infty$. Let us define $\Psi_h^\alpha(z) = U(\Phi_h^\alpha(z))$. Then, there exists $h_0 < \infty$ such that for all $0 < h \leq h_0$ and $\alpha = 0$ or $\frac{1}{\varepsilon}$, one has the estimate*

$$\left(\mathbb{E} \left| \Psi_h^\alpha(z) - U(z) - h(-|\nabla U(z)|^2 + \beta \text{tr}[D^2 U(z)]) - \sqrt{h} \sqrt{2\beta} \nabla U(z) \cdot \xi \right|^2 \right)^{1/2} \leq Ch^{3/2}(1 + \alpha)^{3/2},$$

where ξ is a standard multidimensional normal random variable.

Proof. By applying Taylor expansion to $U(\Phi_h^\alpha(z))$ up to second order with respect to $0 < h \ll h_0$ and using (8) we get for h sufficiently small

$$\begin{aligned} U(\Phi_h^\alpha(z)) &= U(z) + \nabla U(z) (\Phi_h^\alpha(z) - z) + \frac{1}{2} (\Phi_h^\alpha(z) - z)^T D^2 U(z) (\Phi_h^\alpha(z) - z) + o((\Phi_h^\alpha(z) - z)^2) \\ &= U(z) + \nabla U(z) \left(h[-\nabla U(z) + \alpha \tilde{C}(z)] + \sqrt{h} \sqrt{2\beta} \xi + \sqrt{h} \sqrt{\kappa \alpha} \sigma(z) \xi' + I(\alpha, h) \right) \\ &\quad + \frac{1}{2} \left(h[-\nabla U(z) + \alpha \tilde{C}(z)] + \sqrt{h} \sqrt{2\beta} \xi + \sqrt{h} \sqrt{\kappa \alpha} \sigma(z) \xi' + I(\alpha, h) \right)^T D^2 U(z) \times \\ &\quad \times \left(h[-\nabla U(z) + \alpha \tilde{C}(z)] + \sqrt{h} \sqrt{2\beta} \xi + \sqrt{h} \sqrt{\kappa \alpha} \sigma(z) \xi' + I(\alpha, h) \right) + o((\Phi_h^\alpha(z) - z)^2), \end{aligned}$$

where $(\mathbb{E} I^2(\alpha, h))^{1/2} \leq Ch^{3/2}(1 + \alpha)^{3/2}$. Using now the assumptions from Conditions 2 and 3 that $\nabla U(z)C(z) = 0$, $\sigma(z)\sigma^T(z)\nabla U(z) = 0$ and expanding the quadratic term, the previous expression simplifies to

$$U(\Phi_h^\alpha(z)) = U(z) + (h[-|\nabla U(z)|^2 + \beta \text{tr}[D^2 U(z)]] + \sqrt{h} \sqrt{2\beta} \nabla U(z) \xi + \nabla U(z) I(\alpha, h) + R_1(\alpha, h),$$

where $(\mathbb{E} R_1^2(\alpha, h))^{1/2} = o(h^{3/2}(1 + \alpha)^{3/2})$. The latter, essentially concludes the proof of the lemma. \square

For notational convenience, let us define the operator on test functions $f \in \mathcal{C}^2(\mathbb{R})$,

$$(10) \quad \mathcal{L}_Q f(z) = [-|\nabla U(z)|^2 + \beta \text{tr}[D^2 U(z)]] f'(U(z)) + \beta |\nabla U(z)|^2 f''(U(z))$$

Next we have the following lemma for the numerical approximation HMM scheme (9).

Lemma 2. *Let $f \in \mathcal{C}^2(\mathbb{R})$. Then for \bar{Z}_t^ε given by (9) we have, as $\tau < \delta < \tau/\varepsilon$, $\delta, \frac{\tau}{\varepsilon} \downarrow 0$*

$$\mathbb{E} \left(f(U(\bar{Z}_{(n+1)\delta}^\varepsilon)) - f(U(\bar{Z}_{n\delta}^\varepsilon)) \right) = \delta \mathbb{E} \mathcal{L}_Q f(\bar{Z}_{n\delta}^\varepsilon) + O \left(\delta^{3/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \right).$$

Proof. We start by noticing that Lemma 1 implies that

$$U(\bar{Z}_{n\delta+\tau}^\varepsilon) = U(\bar{Z}_{n\delta}^\varepsilon) + \tau (-|\nabla U(\bar{Z}_{n\delta}^\varepsilon)|^2 + \beta \text{tr}[D^2 U(\bar{Z}_{n\delta}^\varepsilon)]) + \sqrt{\tau} \sqrt{2\beta} \nabla U(\bar{Z}_{n\delta}^\varepsilon) \xi_n + R_{2,n},$$

where $(\mathbb{E} R_{2,n}^2)^{1/2} \leq C \left(\frac{\tau}{\varepsilon} \right)^{3/2}$. The last display and smoothness of the test function f , implies that

$$\mathbb{E} f(U(\bar{Z}_{n\delta+\tau}^\varepsilon)) = \mathbb{E} f(U(\bar{Z}_{n\delta}^\varepsilon)) + \tau \mathbb{E} \mathcal{L}_Q f(\bar{Z}_{n\delta}^\varepsilon) + \mathbb{E} R_{2,n},$$

where with some abuse of notation we still denote $R_{2,n}$ the error term which again satisfies $(\mathbb{E} R_{2,n}^2)^{1/2} \leq C \left(\frac{\tau}{\varepsilon} \right)^{3/2}$.

In a similar manner, we also obtain that

$$\mathbb{E}f(U(\bar{Z}_{(n+1)\delta}^\varepsilon)) = \mathbb{E}f(U(\bar{Z}_{n\delta+\tau}^\varepsilon)) + (\delta - \tau)\mathbb{E}\mathcal{L}_Q f(\bar{Z}_{n\delta+\tau}^\varepsilon) + \mathbb{E}R_{3,n}$$

where $(\mathbb{E}R_{3,n}^2)^{1/2} \leq C(\delta - \tau)^{3/2}$.

Hence, we get

$$\begin{aligned} \mathbb{E}\left(f(U(\bar{Z}_{(n+1)\delta}^\varepsilon)) - f(U(\bar{Z}_{n\delta}^\varepsilon))\right) &= \delta\mathbb{E}\mathcal{L}_Q f(\bar{Z}_{n\delta}^\varepsilon) \\ &\quad + (\delta - \tau)\mathbb{E}\left(\mathcal{L}_Q f(\bar{Z}_{n\delta}^\varepsilon) - \mathcal{L}_Q f(\bar{Z}_{n\delta+\tau}^\varepsilon)\right) + \mathbb{E}R_{2,n} + \mathbb{E}R_{3,n} \end{aligned}$$

We further notice that by the regularity of U and f we have

$$\begin{aligned} \left(\mathbb{E}\left(\mathcal{L}_Q f(\bar{Z}_{n\delta}^\varepsilon) - \mathcal{L}_Q f(\bar{Z}_{n\delta+\tau}^\varepsilon)\right)^2\right)^{1/2} &\leq C\left(\mathbb{E}\left|\bar{Z}_{n\delta+\tau}^\varepsilon - \bar{Z}_{n\delta}^\varepsilon\right|^2\right)^{1/2} \leq C(\sqrt{\tau} + \tau/\varepsilon) \\ &\leq C(\sqrt{\delta} + \tau/\varepsilon) \end{aligned}$$

Putting the estimates together we obtain the statement of the lemma. \square

Let us recall now the operator $\mathcal{L}_Q f(z)$ defined by (10) and let us recall the “averaged” generator $\mathcal{L}^Y = \mathcal{L}_1^Y$ defined via (7) (recall that the single edge case is considered at the moment). We want to prove that the process $U(\bar{Z}_{n\delta}^\varepsilon)$ converges in distribution to the process with generator \mathcal{L}^Y .

For this purpose, we may use Theorem 1 of [24, Chapter 2]. By Lemma 2 we have

$$\begin{aligned} (11) \quad &\mathbb{E}\left[f(U(\bar{Z}_{n\delta}^\varepsilon)) - f(U(z)) - \int_0^{n\delta} \mathcal{L}^Y f(U(\bar{Z}_s^\varepsilon)) ds\right] = \\ &= \mathbb{E}\sum_{k=0}^{n-1}\left[f(U(\bar{Z}_{(k+1)\delta}^\varepsilon)) - f(U(\bar{Z}_{k\delta}^\varepsilon)) - \int_{k\delta}^{(k+1)\delta} \mathcal{L}^Y f(U(\bar{Z}_s^\varepsilon)) ds\right] \\ &= \delta\sum_{k=0}^{n-1}\mathbb{E}\left[\mathcal{L}_Q f(\bar{Z}_{k\delta}^\varepsilon) - \mathcal{L}^Y f(U(\bar{Z}_{k\delta}^\varepsilon))\right] + n\delta\mathbb{E}I_0 \\ &= \mathbb{E}\int_0^{n\delta} [\mathcal{L}_Q f(\bar{Z}_s^\varepsilon) - \mathcal{L}^Y f(U(\bar{Z}_s^\varepsilon))] ds + n\delta\mathbb{E}I_0, \end{aligned}$$

where $(\mathbb{E}I_0^2)^{1/2} \leq C\left(\delta^{3/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2}\right)$.

Notice that $\mathcal{L}^Y f(U(z)) = \widehat{\mathcal{L}_Q f}(U(z))$. Essentially, for a nice function $g = \mathcal{L}_Q f$, we need tight estimates for $\mathbb{E}\int_0^{n\delta} [g(\bar{Z}_s^\varepsilon) - \widehat{g}(U(\bar{Z}_s^\varepsilon))] ds$, where we recall that \bar{Z}_s^ε is the approximating process and \widehat{g} is the average on graph as defined in (4). We can write

$$\begin{aligned} \mathbb{E}\int_0^{n\delta} [\mathcal{L}_Q f(\bar{Z}_s^\varepsilon) - \mathcal{L}^Y f(U(\bar{Z}_s^\varepsilon))] ds &= n\delta\left[\frac{1}{n}\sum_{k=0}^{n-1}\left(\mathbb{E}\mathcal{L}_Q f(\bar{Z}_{k\delta}^\varepsilon) - \frac{1}{\tau}\int_{k\tau}^{(k+1)\tau} \mathbb{E}\mathcal{L}_Q f(Z_s^\varepsilon) ds\right)\right] \\ &\quad + n\delta\left[\frac{1}{n\tau}\int_0^{n\tau} \mathbb{E}\mathcal{L}_Q f(Z_s^\varepsilon) ds - \frac{1}{n\tau}\int_0^{n\tau} \mathbb{E}\mathcal{L}^Y f(U(Z_s^\varepsilon)) ds\right] \\ &\quad + n\delta\left[\frac{1}{n}\sum_{k=0}^{n-1}\left(\int_{k\tau}^{(k+1)\tau} \mathbb{E}\mathcal{L}^Y f(U(Z_s^\varepsilon)) ds - \mathbb{E}\mathcal{L}^Y f(U(\bar{Z}_{k\delta}^\varepsilon))\right)\right] \\ &= n\delta[J_1^n + J_2^n + J_3^n] \end{aligned}$$

By the estimate (A.103)-(A.104) of [25] we have that for an unimportant constant $C < \infty$

$$(12) \quad |J_1^n + J_3^n| \leq C \left(\sqrt{\frac{\tau}{\varepsilon}} + \sqrt{n\delta} + n\delta + n \left(\frac{\tau}{\varepsilon} \right)^{3/2} \right) e^{C \frac{n\tau}{\varepsilon}}$$

It remains to treat the term $J_2^n = \frac{1}{n\tau} \int_0^{n\tau} \mathbb{E} \mathcal{L}_Q f(Z_s^\varepsilon) ds - \frac{1}{n\tau} \int_0^{n\tau} \mathbb{E} \mathcal{L}^Y f(U(Z_s^\varepsilon)) ds$. Standard PDE arguments, e.g., [14, Section 3.2], show that for any point $(z, i) \in I_1$ that is not a vertex, and for $g \in \mathcal{C}^{2+\alpha}$, the PDE

$$(13) \quad -\widehat{\mathcal{L}}u(z) = g(z) - \widehat{g}(U(z)), \quad \text{for } z \in d_1(x)$$

has a unique (up to constants) $\mathcal{C}^{2+\alpha'}$ solution with $\alpha' \in (0, \alpha)$. We fix the free constant by setting $\widehat{u}(x, 1) = 0$. Then, the solution $u(z)$ can be written as

$$u(z) = \int_0^\infty \mathbb{E}_z \left[g(\widehat{Z}_s) - \widehat{g}(U(\widehat{Z}_s)) \right] ds.$$

Moreover, there exist a constant $\lambda = \lambda(z, 1) > 0$ such that for $z \in d_1(x)$,

$$|u(z)| \leq \frac{2}{\lambda} \sup_{z \in d_1(x)} |g(z) - \widehat{g}(U(z))|.$$

Notice that in the case that we can take $\kappa = 0$, *i.e.*, when the dynamical system $\dot{z}_t = C(z_t)$ has a unique invariant measure on the connected component $d_1(x)$, then we simply have $\widehat{\mathcal{L}}u(z) = C(z) \nabla u(z)$. If we cannot take $\kappa = 0$, then $\widehat{\mathcal{L}}u(z)$ is given by (6). Applying Itô's formula to u we obtain that

$$\begin{aligned} u(Z_{n\tau}^\varepsilon) &= u(z) + \frac{1}{\varepsilon} \int_0^{n\tau} \widehat{\mathcal{L}}u(Z_s^\varepsilon) ds + \int_0^{n\tau} \mathcal{L}_0 u(Z_s^\varepsilon) ds \\ &\quad + \sqrt{\frac{\kappa}{\varepsilon}} \int_0^{n\delta} \nabla u(Z_s^\varepsilon) \sigma(Z_s^\varepsilon) dW_s^o + \sqrt{2\beta} \int_0^{n\delta} \nabla u(Z_s^\varepsilon) dW_s \end{aligned}$$

Recalling now that u solves (13), we obtain by rearranging the last display and taking expected value

$$\begin{aligned} J_2^n &= \frac{1}{n\tau} \int_0^{n\tau} \mathbb{E} \mathcal{L}_Q f(Z_s^\varepsilon) ds - \frac{1}{n\tau} \int_0^{n\tau} \mathbb{E} \mathcal{L}^Y f(U(Z_s^\varepsilon)) ds \\ &= \frac{\varepsilon}{n\tau} \left[\mathbb{E}(u(Z_{n\tau}^\varepsilon) - u(z)) + \mathbb{E} \int_0^{n\tau} \mathcal{L}_0 u(Z_s^\varepsilon) ds \right], \end{aligned}$$

which then, due to the boundedness of u and its derivatives, gives

$$(14) \quad |J_2^n| \leq C \left(\frac{\varepsilon}{n\tau} + \varepsilon \right).$$

Thus, using estimates (12)-(14), (11) gives

$$\begin{aligned} (15) \quad & \left| \mathbb{E} \left[f(U(\bar{Z}_{n\delta}^\varepsilon)) - f(U(z)) - \int_0^{n\delta} \mathcal{L}^Y f(U(\bar{Z}_s^\varepsilon)) ds \right] \right| \leq \\ & \leq Cn\delta \left[\left(\sqrt{\frac{\tau}{\varepsilon}} + \sqrt{n\delta} + n\delta + n \left(\frac{\tau}{\varepsilon} \right)^{3/2} \right) e^{C \frac{n\tau}{\varepsilon}} + \frac{\varepsilon}{n\tau} + \varepsilon + \delta^{3/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \right] \end{aligned}$$

Choosing now n such that $\sqrt{\frac{n\tau}{\varepsilon}} e^{C\frac{n\tau}{\varepsilon}} \sim \left(\frac{\tau}{\varepsilon\delta}\right)^{1/4}$, and recalling the requirement $\tau < \delta < \frac{\tau}{\varepsilon} \ll 1$ we obtain from (15)

$$\begin{aligned} & \frac{1}{n\delta} \left| \mathbb{E} \left[f(U(\bar{Z}_{n\delta}^\varepsilon)) - f(U(z)) - \int_0^{n\delta} \mathcal{L}^Y f(U(\bar{Z}_s^\varepsilon)) ds \right] \right| \leq \\ & \leq C \left[\left(\frac{\delta\varepsilon}{\tau} \right)^{1/4} + \left(\frac{\delta\varepsilon}{\tau} \right)^{1/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \sqrt{\frac{\delta\varepsilon}{\tau}} + \frac{1}{\log\left(\frac{\tau}{\varepsilon\delta}\right)} + \frac{\tau}{\delta} + \delta^{3/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \right] \\ & \rightarrow 0, \text{ as } \frac{\delta\varepsilon}{\tau}, \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \downarrow 0. \end{aligned}$$

Hence, by Theorem 1 of [24, Chapter 2], we have obtained that $U(\bar{Z}_{n\delta}^\varepsilon)$ converges in distribution to the process Y on the graph (for the moment with just one edge) with generator \mathcal{L}^Y .

Let us next discuss convergence to the invariant measure. Since the invariant measure for the original process Z^ε is the Gibbs measure π , we get that the invariant measure for the process Y on the tree Γ is nothing else but the projection of π on Γ , say μ . In particular for any Borel set $\gamma \subset \Gamma$, we have $\mu(\gamma) = \pi(\Gamma^{-1}(\gamma))$.

Then, from the weak convergence of $Q(\bar{Z}_{n\delta}^\varepsilon)$ to the process Y and the uniform mixing properties of Z_t^ε and Y_t , we get that for any bounded and uniformly Lipschitz test function f that for all $t > 0$

$$\begin{aligned} \lim_{h \downarrow 0} \lim_{\varepsilon, \delta, \frac{\delta\varepsilon}{\tau}, \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta} \downarrow 0} \frac{1}{h} \int_t^{t+h} E_\pi f(\bar{Z}_s^\varepsilon) ds &= \lim_{h \downarrow 0} \lim_{\varepsilon, \delta, \frac{\delta\varepsilon}{\tau}, \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta} \downarrow 0} \frac{1}{h} \int_t^{t+h} E_\pi \hat{f}(Q(\bar{Z}_s^\varepsilon)) ds \\ &= E_\mu \hat{f}(Y_t). \end{aligned} \tag{16}$$

The latter establishes Theorem 6 in the one well case.

5.2. The multi-well case. The goal of this section is to establish that Theorem 6 holds in the general multi-well case, i.e., when $m > 1$. First we need to define certain objects. Let us consider $\theta > 0$ small and for an edge I_i of the graph set

$$I_i^\theta = \{(x, i) \in I_i : \text{dist}((x, i), \partial I_i) > \theta\}$$

and define

$$\bar{\tau}_i = \inf\{t > 0 : Q(\bar{Z}_t^\varepsilon) \notin I_i^\theta\}.$$

Thus, I_i^θ is the interior part of the edge I_i and $\bar{\tau}_i$ is the first exit time of the interior part.

In addition, for $\zeta > 0$ and for a vertex of the graph O_j and a segment $I_i \sim O_j$, let us define the following quantities as in [15, Chapter 8]:

$$\begin{aligned} D_i &= \{z \in E : Q(z) \in I_i^\circ\} \\ D_i(U_1, U_2) &= \{z \in D_i : U_1 < U(z) < U_2\} \\ D_j(\pm\zeta) &= \{z \in E : U(O_j) - \zeta < U(z) < U(O_j) + \zeta\} \\ D(\pm\zeta) &= \bigcup_j D_j(\pm\zeta) \\ C_j &= \{z \in E : Q(z) = O_j\} \\ C_{ji}(\zeta) &= \{z \in D_i : U(z) = U(O_j) \pm \zeta\} \\ C_{ji} &= C_j \cap \partial D_i \\ C_i(U) &= \{z \in \bar{D}_i : U(z) = U\} \end{aligned}$$

Here I_i° denotes the open interior of I_i . Let us then also define the first exit time of the process \bar{Z}_t^ε from $D_j(\pm\zeta)$ as follows

$$\bar{\tau}_j^\varepsilon(\pm\zeta) = \inf\{t > 0 : \bar{Z}_t^\varepsilon \notin D_j(\pm\zeta)\}.$$

Following the proof of [15, Theorem 8.2.2], see also [17], the statement of Theorem 6 will follow if we show that in the limit as $\varepsilon, \delta, \tau \downarrow 0$ the process \bar{Z}_t^ε behaves within a given i well according to the generator \mathcal{L}_i^Y , it spends zero time in exterior and interior vertices and that the probabilistic behavior at the vertices leads to the gluing condition of Definition 4. To be precise, following the proof of [15, Theorem 8.2.2], Theorem 6 follows if we prove Lemma 3, 4, 5 and Lemma 6 below.

Lemma 3. *Let $f \in \mathcal{C}_b^2(\mathbb{R})$ and $\theta > 0$ such that $I_i^\theta \neq \emptyset$ for all $i \in \{1, \dots, m\}$. Assume the conditions of Theorem 6. Then, uniformly in $z \in D_i^\theta = \{z \in E : Q(z) \subset I_i^\theta\}$, we have that*

$$(17) \quad \left| \mathbb{E} \left[f(U(\bar{Z}_{n\delta \wedge \bar{\tau}_i}^\varepsilon)) - f(U(z)) - \int_0^{n\delta \wedge \bar{\tau}_i} \mathcal{L}_i^Y f(U(\bar{Z}_s^\varepsilon)) ds \right] \right| \\ \leq Cn\delta \left(\left(\sqrt{\frac{\tau}{\varepsilon}} + \sqrt{n\delta} + n\delta + n \left(\frac{\tau}{\varepsilon} \right)^{3/2} \right) e^{C\frac{n\tau}{\varepsilon}} + \frac{\varepsilon}{n\tau} + \frac{\tau}{\delta} + \delta^{3/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \right)$$

for some constant $C < \infty$. In particular, choosing n such that $\sqrt{\frac{n\tau}{\varepsilon}} e^{C\frac{n\tau}{\varepsilon}} \sim \left(\frac{\tau}{\varepsilon\delta} \right)^{1/4}$, we obtain that

$$\frac{1}{n\delta} \left| \mathbb{E} \left[f(U(\bar{Z}_{n\delta \wedge \bar{\tau}_i}^\varepsilon)) - f(U(z)) - \int_0^{n\delta \wedge \bar{\tau}_i} \mathcal{L}_i^Y f(U(\bar{Z}_s^\varepsilon)) ds \right] \right| \leq \\ \leq C \left[\left(\frac{\delta\varepsilon}{\tau} \right)^{1/4} + \left(\frac{\delta\varepsilon}{\tau} \right)^{1/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \sqrt{\frac{\delta\varepsilon}{\tau}} + \frac{1}{\log\left(\frac{\tau}{\varepsilon\delta}\right)} + \frac{\tau}{\delta} + \delta^{3/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \right] \\ \rightarrow 0, \text{ as } \frac{\delta\varepsilon}{\tau}, \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \downarrow 0.$$

Lemma 3 follows directly by the arguments of Section 5.1. In particular Lemma 3 implies that if $\delta, \frac{\delta\varepsilon}{\tau}, \frac{\tau}{\varepsilon}, \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \downarrow 0$, then the process $Q(\bar{Z}_{n\delta \wedge \bar{\tau}_i}^\varepsilon)$ converges in distribution, within edge I_i , to the process with generator \mathcal{L}_i^Y as defined by (7).

Lemma 4. *Let O_j be an exterior vertex of the graph Γ . Assume the conditions of Theorem 6 and that $U \in \mathcal{C}_b^2$. Then, there exists $\zeta_0 > 0$, such that for all $0 < \zeta \leq \zeta_0$ and for all $z \in D_j(\pm\zeta)$, there exists a constant $C < \infty$ such that*

$$\mathbb{E}_z \bar{\tau}_j^\varepsilon(\pm\zeta) \leq C(\zeta + \delta + (\tau/\varepsilon)^{3/2}).$$

In other words, for every $\kappa > 0$ and for $0 < \max\{\delta, (\tau/\varepsilon)^{3/2}\} < \zeta \leq \zeta_0$ sufficiently small, we have that

$$\mathbb{E}_z \bar{\tau}_j^\varepsilon(\pm\zeta) \leq \kappa.$$

Lemma 5. *Let O_j be an interior vertex of the graph Γ . Assume the conditions of Theorem 6 and that $U \in \mathcal{C}_b^2$. Then, there exists $\zeta_0 > \varepsilon^\alpha$ for some exponent $\alpha > 0$, such that for all $0 < \varepsilon^\alpha < \zeta \leq \zeta_0$ and for all $z \in D_j(\pm\zeta)$*

$$\mathbb{E}_z \bar{\tau}_j^\varepsilon(\pm\zeta) \leq C\zeta^2 |\ln \zeta|.$$

In other words for every $\kappa > 0$, there exists $\zeta_0 > \varepsilon^\alpha$ for some exponent $\alpha > 0$, such that for all $0 < \varepsilon^\alpha < \zeta \leq \zeta_0$ and for all $z \in D_j(\pm\zeta)$

$$\mathbb{E}_z \bar{\tau}_j^\varepsilon(\pm\zeta) \leq \kappa\zeta.$$

Lemma 6. For $I_i \sim O_j$ define b_{ji} as in Definition 4 and set $p_{ji} = \frac{b_{ji}}{\sum_{i: I_i \sim O_j} b_{ji}}$. Assume the conditions of Theorem 6 and that $U \in \mathcal{C}_b^2$. Then, for every $\kappa > 0$, there exists $\zeta_0 > \max\{\delta, (\tau/\varepsilon)^{3/2}\}$, such that for all $0 < \max\{\delta, (\tau/\varepsilon)^{3/2}\} < \zeta \leq \zeta_0$ there exists $\zeta'_0 = \zeta'_0(\zeta)$ such that for all sufficiently small $\varepsilon, \delta, \tau$

$$\left| \mathbb{P}_z \left(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon \notin \overline{\partial D_j(\pm\zeta)} \cap I_i^\circ \right) - p_{ji} \right| < \kappa.$$

for all $z \in \overline{D_j(\pm\zeta'_0)}$.

Lemmas 4, 5 and 6 follow as Lemmas 3.4, 3.5 and 3.6 in [15, Chapter 8]. The main difference between our situation and that of [15] is that we are working with the discrete approximation, which implies that we need information on the error bounds in terms of the parameters $\delta, \varepsilon, \tau$, as it was also the case for Lemma 3. Given that the method of the proof is similar to the corresponding proofs of [15], we do not repeat all the details here.

The principle idea is that Lemma 3 controls the behavior within each branch of the tree, whereas Lemmas 4, 5 allow us to conclude that the approximating process spends in the limit zero time on exterior and interior vertices respectively (equivalently it spends zero time in the neighborhood of stable and unstable points of the dynamical system). Then, Lemma 6 characterizes the splitting probability in each interior vertex concluding the description of the limiting Markov process.

In order to demonstrate the differences with the corresponding proofs of [15] and to see the role of the discrete approximation, we demonstrate the proofs of these lemmas emphasizing the differences.

Proof of Lemma 4. Let us assume that $U(z_j)$ is a local minimum of U . It is clear that the following relation should hold

$$|\mathbb{E}U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(z_j)| \geq \zeta.$$

Let us define

$$k_1 = \max\{k \in \mathbb{N} : (k\delta + \tau) \vee k\delta \leq \bar{\tau}_j^\varepsilon(\pm\zeta)\}.$$

Let us assume that k_1 is such that $(k_1\delta + \tau) \leq \bar{\tau}_j^\varepsilon(\pm\zeta)$. The approach is the same if k_1 is such that $k_1\delta \leq \bar{\tau}_j^\varepsilon(\pm\zeta)$. By adding and subtracting terms of the form $U(\bar{Z}_{m\delta}^\varepsilon)$ for $m = 0, 1, \dots, k_1$ we get

$$\begin{aligned} U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) &= \left[U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) \right] + U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) \\ &= \left[U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) \right] + \left[U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) - U(\bar{Z}_{k_1\delta}^\varepsilon) \right] + U(\bar{Z}_{k_1\delta}^\varepsilon) \\ &= \left[U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) \right] + \left[U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) - U(\bar{Z}_{k_1\delta}^\varepsilon) \right] + \sum_{m=1}^{k_1} \left[U(\bar{Z}_{m\delta}^\varepsilon) - U(\bar{Z}_{(m-1)\delta}^\varepsilon) \right] + U(z). \end{aligned}$$

Taking expected value, Lemma 2 (with the test function $f(u) = u$) implies that for $\delta, \tau/\varepsilon$ sufficiently small

$$\begin{aligned} \mathbb{E}U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) &= \mathbb{E} \left[U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) \right] + \mathbb{E} \left[U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) - U(\bar{Z}_{k_1\delta}^\varepsilon) \right] \\ &\quad + \delta \mathbb{E} \sum_{m=1}^{k_1} \left[\mathcal{L}_0 U(\bar{Z}_{(m-1)\delta}^\varepsilon) + O \left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \right) \right] + U(z). \end{aligned}$$

Next we notice that up to an unimportant multiplicative constant $|\mathbb{E}U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(z)| < \zeta + \delta$ and that the non-degeneracy of U implies that for every $z \in D_j(\pm\zeta)$ there is a constant $C_0 > 0$, such that $\mathcal{L}_0 U(z) > C_0$. Hence, we have obtained

$$\begin{aligned} \zeta + \delta &> \mathbb{E} \left[U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) \right] + \mathbb{E} \left[U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) - U(\bar{Z}_{k_1\delta}^\varepsilon) \right] \\ &\quad + C\mathbb{E}\bar{\tau}_j^\varepsilon(\pm\zeta) \left(1 + O \left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \right) \right) + C\mathbb{E}(\delta k_1 - \bar{\tau}_j^\varepsilon(\pm\zeta)) \left(1 + O \left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \right) \right) \\ &\geq \mathbb{E} \left[U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) \right] + \mathbb{E} \left[U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) - U(\bar{Z}_{k_1\delta}^\varepsilon) \right] \\ &\quad + C_0\mathbb{E}\bar{\tau}_j^\varepsilon(\pm\zeta) \left(1 + O \left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \right) \right) - \delta \left(1 + O \left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \right) \right). \end{aligned}$$

The latter inequality follows since by the definition of k_1 we have that $|\delta k_1 - \bar{\tau}_j^\varepsilon(\pm\zeta)| < \delta$. Rearranging the latter expression, we obtain for some unimportant constants $0 < C_i < \infty$

$$\begin{aligned} \mathbb{E}\bar{\tau}_j^\varepsilon(\pm\zeta) &\leq C_1 \frac{\zeta + 2\delta + \delta^{3/2} + (\frac{\tau}{\varepsilon})^{3/2} + \left| \mathbb{E} \left[U(\bar{Z}_{\bar{\tau}_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) \right] \right| + \left| \mathbb{E} \left[U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) - U(\bar{Z}_{k_1\delta}^\varepsilon) \right] \right|}{1 + O \left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \right)} \\ &\leq C_2 \frac{\zeta + 2\delta + \delta^{3/2} + (\frac{\tau}{\varepsilon})^{3/2} + (\delta + \tau) + (\delta^{3/2} + (\tau/\varepsilon)^{3/2})}{1 + O \left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \frac{1}{\delta} \right)} \end{aligned}$$

which implies that for $0 < \tau < \delta < \frac{\tau}{\varepsilon} \ll 1$ and $(\frac{\tau}{\varepsilon})^{3/2} \frac{1}{\delta} \downarrow 0$, we get

$$\mathbb{E}\bar{\tau}_j^\varepsilon(\pm\zeta) \leq C_3 \left(\zeta + \delta + \left(\frac{\tau}{\varepsilon} \right)^{3/2} \right)$$

or, in other words if we choose $\zeta > \max\{\delta, (\frac{\tau}{\varepsilon})^{3/2}\}$, we indeed obtain that

$$\mathbb{E}\bar{\tau}_j^\varepsilon(\pm\zeta) \leq C_4 \zeta$$

from which the statement of the lemma follows. \square

Proof of Lemma 5. We start with the following usage of Lemma 2,

$$\begin{aligned} U(\bar{Z}_{n\delta}^\varepsilon) &= \sum_{k=0}^{n-1} \left[U(\bar{Z}_{(k+1)\delta}^\varepsilon) - U(\bar{Z}_{k\delta}^\varepsilon) \right] + U(z) \\ &= \sum_{k=0}^{n-1} \left[U(\bar{Z}_{(k+1)\delta}^\varepsilon) - U(\bar{Z}_{k\delta+\tau}^\varepsilon) \right] + \sum_{k=0}^{n-1} \left[U(\bar{Z}_{k\delta+\tau}^\varepsilon) - U(\bar{Z}_{k\delta}^\varepsilon) \right] + U(z) \\ &= \sum_{k=0}^{n-1} \left[(\delta - \tau)\mathcal{L}_0 U(\bar{Z}_{k\delta+\tau}^\varepsilon) + \tau\mathcal{L}_0 U(\bar{Z}_{k\delta}^\varepsilon) \right] + \\ &\quad + \sqrt{2\beta} \sum_{k=0}^{n-1} \left[\sqrt{\delta - \tau} \nabla U(\bar{Z}_{k\delta+\tau}^\varepsilon) \xi_k' + \sqrt{\tau} \nabla U(\bar{Z}_{k\delta}^\varepsilon) \xi_k \right] + \sum_{k=0}^{n-1} [R_{2,k} + R_{3,k}] + U(z), \end{aligned}$$

where ξ_k', ξ_k are independent standard normal random variables and $R_{2,k}, R_{3,k}$ are as in the proof of Lemma 2. Using the independence of the involved normal random variables, we can then write that in distribution

$$(18) \quad U(\bar{Z}_{n\delta}^\varepsilon) = \sum_{k=0}^{n-1} I_{1,k}^{\delta,\tau} + N \left(0, \sum_{k=0}^{n-1} I_{2,k}^{\delta,\tau} \right) + \sum_{k=0}^{n-1} [R_{2,k} + R_{3,k}] + U(z)$$

where

$$\begin{aligned} I_{1,k}^{\delta,\tau} &= [(\delta - \tau)(\mathcal{L}_0 U(\bar{Z}_{k\delta+\tau}^\varepsilon) - \mathcal{L}_0 U(\bar{Z}_{k\delta}^\varepsilon)) + \delta \mathcal{L}_0 U(\bar{Z}_{k\delta}^\varepsilon)], \\ I_{2,k}^{\delta,\tau} &= 2\beta [(\delta - \tau)|\nabla U(\bar{Z}_{k\delta+\tau}^\varepsilon)|^2 + \tau|\nabla U(\bar{Z}_{k\delta}^\varepsilon)|^2] \end{aligned}$$

and $N\left(0, \sum_{k=0}^{n-1} I_{2,k}^{\delta,\tau}\right)$ represents a normal random variable with mean zero and variance $\sum_{k=0}^{n-1} I_{2,k}^{\delta,\tau}$.

Let us recall now that

$$\left(\mathbb{E} \left(\sum_{k=0}^{n-1} [R_{2,k} + R_{3,k}] \right)^2 \right)^{1/2} = O(n\delta^{3/2} + n(\tau/\varepsilon)^{3/2}).$$

Moreover, it is clear that if $\sum_{k=0}^{n-1} I_{2,k}^{\delta,\tau} \geq 9\zeta^2$, then

$$\max_{1 \leq j \leq n} \left| N \left(0, \sum_{k=0}^{j-1} I_{2,k}^{\delta,\tau} \right) \right| \geq |N(0, 9\zeta^2)|.$$

At the same time we have that $\bar{\tau}_j^\varepsilon(\pm\zeta)$ is less or equal to the time when the random variable $|U(\bar{Z}_{n\delta}^\varepsilon) - U(z)|$ reaches the level 2ζ . This happens if the term $\sum_{k=0}^{n-1} [I_{1,k}^{\delta,\tau} + R_{2,k} + R_{3,k}]$ is small in absolute value, while the term $N\left(0, 2\beta \sum_{k=0}^{n-1} (\delta - \tau)|\nabla U(\bar{Z}_{k\delta+\tau}^\varepsilon)|^2 + \tau|\nabla U(\bar{Z}_{k\delta}^\varepsilon)|^2\right)$ is large. In other words we have the inclusion

$$(19) \quad \left\{ \sum_{k=0}^{n-1} [I_{1,k}^{\delta,\tau} + R_{2,k} + R_{3,k}] < \zeta, N \left(0, \sum_{k=0}^{n-1} I_{2,k}^{\delta,\tau} \right) > 3\zeta \right\} \subseteq \{ \bar{\tau}_j^\varepsilon(\pm\zeta) < n\delta \}.$$

We also have

$$\begin{aligned} \{ \bar{\tau}_j^\varepsilon(\pm\zeta) \geq n\delta \} &\subseteq \left\{ \bar{\tau}_j^\varepsilon(\pm\zeta) \geq n\delta, \sum_{k=0}^{n-1} I_{2,k}^{\delta,\tau} < 9\zeta^2 \right\} \cup \\ &\cup \left\{ \bar{\tau}_j^\varepsilon(\pm\zeta) \geq n\delta, \sum_{k=0}^{n-1} I_{2,k}^{\delta,\tau} \geq 9\zeta^2, |N(0, 9\zeta^2)| \geq 3\zeta \right\} \cup \\ &\cup \left\{ \bar{\tau}_j^\varepsilon(\pm\zeta) \geq n\delta, \sum_{k=0}^{n-1} I_{2,k}^{\delta,\tau} \geq 9\zeta^2, |N(0, 9\zeta^2)| < 3\zeta \right\}. \end{aligned}$$

Choose now $n\delta$ such that for the given ζ we have $n\delta < \zeta$ and in particular that

$$\sum_{k=0}^{n-1} [I_{1,k}^{\delta,\tau} + R_{2,k} + R_{3,k}] < \zeta$$

for all trajectories \bar{Z}^ε for which $\bar{\tau}_j^\varepsilon(\pm\zeta) \geq n\delta$. Then, by (19), the second inclusion in the last display cannot hold. Thus we have

$$\begin{aligned} \mathbb{P}(\bar{\tau}_j^\varepsilon(\pm\zeta) \geq n\delta) &\leq \mathbb{P} \left(\bar{\tau}_j^\varepsilon(\pm\zeta) \geq n\delta, \sum_{k=0}^{n-1} I_{2,k}^{\delta,\tau} < 9\zeta^2 \right) + \mathbb{P}(|N(0, 9\zeta^2)| < 3\zeta) \\ &= \mathbb{P} \left(\bar{\tau}_j^\varepsilon(\pm\zeta) \geq n\delta, \sum_{k=0}^{n-1} I_{2,k}^{\delta,\tau} < 9\zeta^2 \right) + 0.6826. \end{aligned}$$

Recall that we have chosen n such that $n\delta < \zeta$ and in particular that $\sum_{k=0}^{n-1} [I_{1,k}^{\delta,\tau} + R_{2,k} + R_{3,k}] < \zeta$. To be precise, the last requirement is that up to a deterministic constant $n(\delta + \delta^{3/2} + (\tau/\varepsilon)^{3/2}) <$

ζ . Let us enforce that by requiring that up to an appropriate deterministic constants $\zeta^2 < n(\delta + \delta^{3/2} + (\tau/\varepsilon)^{3/2}) < \zeta$. In particular, we can take $n(\delta + \delta^{3/2} + (\tau/\varepsilon)^{3/2})$ to be of the order of $\zeta^2 |\ln \varepsilon|$ such that $\zeta |\ln \varepsilon| \rightarrow 0$. Then, the probability of the first term in the right hand side of the last display can be made as small as we want, say less than 0.10.

Hence, we have obtained that with the particular choices for n and for sufficiently small $\varepsilon, \delta, \tau$ and ζ such that $\zeta^2 < n(\delta + \delta^{3/2} + (\tau/\varepsilon)^{3/2}) < \zeta$ and $n(\delta + \delta^{3/2} + (\tau/\varepsilon)^{3/2})$ to be of the order of $\zeta^2 |\ln \varepsilon|$, we have that

$$\mathbb{P}(\bar{\tau}_j^\varepsilon(\pm\zeta) \geq n\delta) \leq 0.8$$

Then, by Markov property we obtain that $\mathbb{P}(\bar{\tau}_j^\varepsilon(\pm\zeta) \geq Nn\delta) \leq 0.8^N$, which then implies that up to deterministic constants

$$\mathbb{E}_z \bar{\tau}_j^\varepsilon(\pm\zeta) \leq \frac{n\delta}{1-0.8} \leq \zeta^2 |\ln \varepsilon| \leq \zeta^2 |\ln \zeta|,$$

if $\zeta \geq \varepsilon^\alpha$ for some exponent $\alpha > 0$. This concludes the proof of the lemma. \square

Proof of Lemma 6. Using [15, Lemma 8.6.2] for the discrete approximation \bar{Z}_t^ε we have

$$\lim_{\varepsilon, \delta, \frac{\tau}{\varepsilon} \downarrow 0} \max_{x_1, x_2 \in C_i(U)} \max_{f: \|f\| \leq 1} \left| \mathbb{E}_{x_1} f(\bar{Z}_{\bar{\tau}^\varepsilon(U_1, U_2)}^\varepsilon) - E_{x_2} f(\bar{Z}_{\bar{\tau}^\varepsilon(U_1, U_2)}^\varepsilon) \right| = 0,$$

where f is defined on $\partial D_i(U_1, U_2)$ and $\bar{\tau}^\varepsilon(U_1, U_2)$ is the first time of exit of the process \bar{Z}^ε from the branch I_i from either of the two sides $U_1 < U_2$. Then, by Markov property, as in [15, Lemma 8.6.3], we get that

$$(20) \quad \lim_{\varepsilon, \delta, \frac{\tau}{\varepsilon} \downarrow 0} \max_{x_1, x_2 \in C_{ji}(\zeta')} |F^\varepsilon(x_1) - F^\varepsilon(x_2)| = 0,$$

where for $\zeta' < \zeta$, $F^\varepsilon(x) = \mathbb{P}(\bar{Z}_{\bar{\tau}^\varepsilon(\pm\zeta)}^\varepsilon \notin \overline{\partial D_j(\pm\zeta)} \cap I_i^\circ)$. The next thing to prove is that for every $\zeta > 0, \kappa > 0$ there exists $0 < \zeta' < \zeta$ such that for every $\varepsilon, \delta, \frac{\tau}{\varepsilon}$ sufficiently small

$$\max_{x_1, x_2 \in \bar{D}_j(\pm\zeta')} |F^\varepsilon(x_1) - F^\varepsilon(x_2)| < \kappa.$$

For edges $I_i \sim O_j$ let us set $f(x) = 1_{x \notin \overline{\partial D_j(\pm\zeta)} \cap I_i^\circ}$. There are exactly three regions corresponding to $I_{i_0}^\circ, I_{i_1}^\circ, I_{i_2}^\circ$ that are separated by the separatrix C_j . The region corresponding to $I_{i_0}^\circ$ adjoins the whole curve C_j , whereas $I_{i_1}^\circ, I_{i_2}^\circ$ adjoins only part of it. In particular we have that $C_{ji_0} = C_{ji_1} \cup C_{ji_2}$. Then, as in the proof of [15, Lemma 8.3.6], it can be shown that

$$(21) \quad \begin{aligned} |F^\varepsilon(x_1) - F^\varepsilon(x_2)| &\leq \left[\sup_{x \in \partial D_j(\pm\zeta)} f(x) - \inf_{x \in \partial D_j(\pm\zeta)} f(x) \right] \\ &\quad \times \max \left\{ \mathbb{P}_{x_m} \left(\bar{Z}_{\bar{\tau}}^\varepsilon \notin (\overline{\partial D_j(\pm\zeta)} \cap I_{i_1}^\circ) \cup (\overline{\partial D_j(\pm\zeta)} \cap I_{i_2}^\circ) \right) : m = 1, 2 \right\} \\ &\quad + \left[\sup_{x \in C_{ji_0}(\zeta')} F^\varepsilon(x) - \inf_{x \in C_{ji_0}(\zeta')} F^\varepsilon(x) \right], \end{aligned}$$

where $\bar{\tau}$ is the first time that the discrete approximation process \bar{Z}_t^ε exits $(\overline{\partial D_j(\pm\zeta')} \cap I_{i_0}^\circ)$ or $(\overline{\partial D_j(\pm\zeta)} \cap I_{i_1}^\circ) \cup (\overline{\partial D_j(\pm\zeta)} \cap I_{i_2}^\circ)$. Clearly, we have that $\bar{\tau} \leq \bar{\tau}^\varepsilon(\pm\zeta)$.

By (20), the second additive term in (21) is arbitrarily small for sufficiently small $\varepsilon, \delta, \tau/\varepsilon$. So, it remains to estimate $P_x^\varepsilon \doteq \mathbb{P}_x(\bar{Z}_{\bar{\tau}}^\varepsilon \notin (\overline{\partial D_j(\pm\zeta)} \cap I_{i_1}^\circ) \cup (\overline{\partial D_j(\pm\zeta)} \cap I_{i_2}^\circ))$. As in the proof of Lemma

4 for an appropriate integer k_1 and for $\delta, \tau/\varepsilon$ sufficiently small

$$(22) \quad \begin{aligned} \mathbb{E}U(\bar{Z}_\tau^\varepsilon) &= \mathbb{E} [U(\bar{Z}_\tau^\varepsilon) - U(\bar{Z}_{k_1\delta+\tau}^\varepsilon)] + \mathbb{E} [U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) - U(\bar{Z}_{k_1\delta}^\varepsilon)] \\ &\quad + \delta \mathbb{E} \sum_{m=1}^{k_1} \left[\mathcal{L}_0 U(\bar{Z}_{(m-1)\delta}^\varepsilon) + O\left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta}\right) \right] + U(z). \end{aligned}$$

Notice now that $U(\bar{Z}_\tau^\varepsilon)$ is either greater or equal than $U(O_j) \pm \zeta'$ on $(\overline{\partial D_j(\pm\zeta')} \cap I_{i_0}^\circ)$ or it is greater or equal than $U(O_j) \mp \zeta$ on $(\overline{\partial D_j(\pm\zeta)} \cap I_{i_1}^\circ) \cup (\overline{\partial D_j(\pm\zeta)} \cap I_{i_2}^\circ)$. The latter implies that

$$\mathbb{E}U(\bar{Z}_\tau^\varepsilon) \geq (U(O_j) \pm \zeta')(1 - P_x^\varepsilon) + (U(O_j) \mp \zeta)P_x^\varepsilon$$

The latter and (22) imply that up to deterministic constants that do not depend on the small parameters of the problem

$$\begin{aligned} (\zeta + \zeta')P_x^\varepsilon &\leq \zeta' + |U(O_j) - U(z)| + |\mathbb{E} [U(\bar{Z}_{\tau_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(\bar{Z}_{k_1\delta+\tau}^\varepsilon)]| + |\mathbb{E} [U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) - U(\bar{Z}_{k_1\delta}^\varepsilon)]| \\ &\quad + \left| \delta \mathbb{E} \sum_{m=1}^{k_1} \left[\mathcal{L}_0 U(\bar{Z}_{(m-1)\delta}^\varepsilon) + O\left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta}\right) \right] \right| \\ &\leq 2\zeta' + |\mathbb{E} [U(\bar{Z}_{\tau_j^\varepsilon(\pm\zeta)}^\varepsilon) - U(\bar{Z}_{k_1\delta+\tau}^\varepsilon)]| + |\mathbb{E} [U(\bar{Z}_{k_1\delta+\tau}^\varepsilon) - U(\bar{Z}_{k_1\delta}^\varepsilon)]| \\ &\quad + \mathbb{E}\bar{\tau} \left(1 + O\left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta}\right) \right) + |\mathbb{E}(\delta k_1 - \bar{\tau})| \left(1 + O\left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta}\right) \right) \\ &\leq 2\zeta' + \delta + \tau + \delta^{3/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2} + \mathbb{E}\bar{\tau} \left(1 + O\left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta}\right) \right) \\ &\leq 2\zeta' + \delta + \tau + \delta^{3/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2} + \mathbb{E}\bar{\tau}^\varepsilon(\pm\zeta) \left(1 + O\left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta}\right) \right) \\ &\leq 2\zeta' + \delta + \tau + \delta^{3/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2} + \zeta^2 |\ln \zeta| \left(1 + O\left(\delta^{1/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta}\right) \right) \end{aligned}$$

where for the last line we used Lemma 5. Therefore, we have obtained that for sufficiently small $\tau < \delta \ll 1$ such that $\tau/\varepsilon \downarrow 0$

$$\begin{aligned} \mathbb{P}_x \left(\bar{Z}_\tau^\varepsilon \notin (\overline{\partial D_j(\pm\zeta)} \cap I_{i_1}^\circ) \cup (\overline{\partial D_j(\pm\zeta)} \cap I_{i_2}^\circ) \right) &\leq \frac{2\zeta'}{\zeta} + \zeta |\ln \zeta| + \frac{\delta + \tau + \delta^{3/2} + \left(\frac{\tau}{\varepsilon}\right)^{3/2}}{\zeta + \zeta'} \\ &\leq \frac{2\zeta'}{\zeta} + \zeta |\ln \zeta| + \frac{\delta + \left(\frac{\tau}{\varepsilon}\right)^{3/2}}{\zeta + \zeta'} \end{aligned}$$

The right hand side of the last display can be made arbitrarily small, if we choose $\zeta' < \zeta$ small but such that $\frac{\delta + \left(\frac{\tau}{\varepsilon}\right)^{3/2}}{\zeta + \zeta'} \downarrow 0$. This means that $\zeta' < \zeta$ should be chosen small, but greater than $\delta + \left(\frac{\tau}{\varepsilon}\right)^{3/2}$. Hence, under this condition, we get that $\mathbb{P}_x \left(\bar{Z}_\tau^\varepsilon \notin (\overline{\partial D_j(\pm\zeta)} \cap I_i^\circ) \right)$ has approximately the same value for all $x \in \bar{D}_j(\pm\zeta')$ when $\tau < \delta < \frac{\tau}{\varepsilon} \ll 1$ and $\left(\frac{\tau}{\varepsilon}\right)^{3/2} \frac{1}{\delta} \ll 1$. Then, it remains to show that, in the limit, this value is actually equal to p_{ji} . This part of the proof however follows very closely the corresponding part of the proof of [15, Lemma 3.6] for Z^ε when $\delta, \tau/\varepsilon$ are sufficiently small and it will not be repeated here. This concludes the proof of the lemma. \square

REFERENCES

- [1] A. Abdulle and S. Cirilli, *S-ROCK: Chebyshev methods for stiff stochastic differential equations*, SIAM J. Sci. Comput. **30** (2008), 997–1014.

- [2] A. Abdulle, D. Cohen, G. Vilmart, and K. C. Zygalakis, *High weak order methods for stochastic differential equations based on modified equations*, SIAM J. Sci. Comput. **34** (2012), A1800–A1823.
- [3] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden, *The heterogeneous multiscale method*, Acta Numerica (2012), 1–87.
- [4] A. Abdulle and T. Li, *S-ROCK methods for stiff Ito SDEs*, Commun. Math. Sci. **6** (2008), 845–868.
- [5] M. Brin and M. I. Freidlin, *On stochastic behavior of perturbed Hamiltonian systems*, Ergodic Theory and Dynamical Systems **20** (2000), 55–76.
- [6] K. Burrage and T. Tian, *Stiffly accurate Runge-Kutta methods for stiff stochastic differential equations*, Comput. Phys. Commun. **142** (2001), 186–190.
- [7] A. Duncan, T. Lelièvre, and G.A. Pavliotis, *Variance reduction using nonreversible Langevin samplers*, J. Stat. Phys. **163** (2016), 457–491.
- [8] W. E, *Principles of multiscale modeling*, Cambridge University Press, Cambridge, 2011.
- [9] W. E and B. Engquist, *The heterogeneous multi-scale methods*, Commun. Math. Sci. **1** (2003), 87–133.
- [10] W. E, B. Engquist, X. Li, and W. Ren, *Heterogeneous multiscale methods: A review*, Communications in Computational Physics **2** (2007), 367–450.
- [11] W. E, D. Liu, and E. Vanden-Eijnden, *Analysis of multiscale methods for stochastic differential equations*, Comm. Pure App. Math. **58** (2005), 1544–1585.
- [12] W. E and J. Lu, *Seamless multiscale modeling via dynamics on fiber bundles*, Commun. Math. Sci. **5** (2007), 649–663.
- [13] W. E, W. Ren, and E. Vanden-Eijnden, *A general strategy for designing seamless multiscale methods*, J. Comput. Phys. **228** (2009), 5437–5453.
- [14] M. I. Freidlin and M. Weber, *Random perturbations of dynamical systems and diffusion processes with conservation laws*, Probability Theory and Related Fields **128** (2004), 441–466.
- [15] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*, 2nd ed., Springer-Verlag, New York, 1988.
- [16] M. I. Freidlin and A. D. Wentzell, *Diffusion processes on graphs and the averaging principle*, Annals of Probability **21** (1993), 2215–2245.
- [17] M. I. Freidlin and A. D. Wentzell, *Random perturbations of Hamiltonian systems*, Memoirs of the American Mathematical Society **109** (1994), no. 523.
- [18] D. Givon, I. G. Kevrekidis, and R. Kupferman, *Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems*, Commun. Math. Sci. **4** (2006), 707–729.
- [19] C.-R. Hwang, S.Y. Hwang-Ma, and S.-J. Sheu, *Accelerating diffusions*, The Annals of Applied Probability **15** (2005), 1433–1444.
- [20] G.A. Pavliotis and A.M. Stuart, *Multiscale methods: Averaging and homogenization*, Springer-Verlag, New York, 2008.
- [21] L. Rey-Bellet and K. Spiliopoulos, *Irreversible Langevin samplers and variance reduction: a large deviation approach*, Nonlinearity **28** (2015), 2081–2103.
- [22] L. Rey-Bellet and K. Spiliopoulos, *Variance reduction for irreversible Langevin samplers and diffusion on graphs*, Electronic Communications in Probability **20** (2015), 1–16.
- [23] G. Roberts and R. Tweedie, *Exponential convergence of Langevin distributions and their discrete approximations*, Bernoulli **2** (1996), 341–363.
- [24] A. Skorokhod, *Asymptotic methods in the theory of stochastic differential equations*, AMS Translation of Mathematical Monographs, vol. 78, American Mathematical Society, Providence, RI, 1989.
- [25] M. Tao, H. Owhadi, and J. E. Marsden, *Nonintrusive and structure preserving multiscale integration of stiff ODEs, SDEs, and Hamiltonian systems with hidden slow dynamics via flow averaging*, Multiscale Modeling and Simulation **8** (2010), 1269–1324.
- [26] E. Vanden-Eijnden, *Numerical techniques for multiscale dynamical systems with stochastic effects*, Commun. Math. Sci. **1** (2003), 385–391.
- [27] E. Vanden-Eijnden, *On HMM-like integrators and projective integration methods for systems with multiple time scales*, Commun. Math. Sci. **5** (2007), 495–505.

DEPARTMENT OF MATHEMATICS, DEPARTMENT OF PHYSICS, AND DEPARTMENT OF CHEMISTRY, DUKE UNIVERSITY, BOX 90320, DURHAM NC 27708, USA

E-mail address: `jianfeng@math.duke.edu`

DEPARTMENT OF MATHEMATICS AND STATISTICS, BOSTON UNIVERSITY, BOSTON MA 02446, USA

E-mail address: `kspiliop@math.bu.edu`